DOCUMENT RESUME

ED 230 198 IR 010 721

AUTHOR	Buckley, Joseph J.
TITLE	A Computerized Information Retrieval System for the Technical Assistance Center Clearinghouse: A Feasibility Study.
INSTITUTION	California Univ., Santa Barbara.
SPONS AGENCY	Office of Education (DHEW), Washington, D.C.
PUB DATE	Oct 79
CONTRACT	300-79-0487
NOTE	16p.
PUB TYPE	Reports - Evaluative/Feasibility (142)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*Computer Programs; Cost Effectiveness; *Databases; *Evaluation Criteria; Feasibility Studies; Information Needs; *Information Networks; *Information Retrieval; Technical Assistance
IDENTIFIERS	Database Management Systems; DIALOG; Stanford Public Information Retrieval System; *Technical Assistance Centers

ABSTRACT

This report first analyzes the need for a Technical Assistance Center (TAC) Clearinghouse, which would provide a range of services including question answering for state and local education agencies, workshop preparation assistance, and literature surveys. Addressing the question of feasibility of a computerized information retrieval system for the TAC, the second section discusses the capabilities and use of such a system, lists costs, examines the available options, and concludes that an information retrieval system would be both appropriate and cost effective. Software packages are discussed in terms of five evaluation criteria: simplicity, flexibility, comprehensiveness, efficiency, and cost. A comparison based on these criteria is made of a software package available in-house (Lockheed's DIALOG Private File Service) and the Stanford Public Information Retrieval System (SPIRES), a locally-available, commercial product. Additional criteria for use in evaluating the overall effectiveness of an information retrieval system are listed. Conclusions are presented which support the establishment of a TAC clearinghouse and the use of SPIRES is recommended. Seven references are listed. (LMM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED230198

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

A COMPUTERIZED INFORMATION RETRIEVAL SYSTEM FOR THE TECHNICAL ASSISTANCE CENTER CLEARINGHOUSE: A FEASIBILITY STUDY

by

Joseph J. Buckley

American Institutes for Research in the Behavioral Sciences
PO Box 1113
1791 Arastradero Rd.
Palo Alto, CA 94302

October 1979

IR010721

TABLE OF CONTENTS

Page

INTRODUCTION 1

I. Background Information 2

II. A Computerized, Bibliographic Information Retrieval System 4

III. Software Packages 7

- Lockheed's DIALOG Private File Service 7
- The Stanford Public Information Retrieval System 8

IV. Evaluation 10

V. Conclusions 12

REFERENCES 13

A COMPUTERIZED INFORMATION RETRIEVAL SYSTEM FOR THE TECHNICAL ASSISTANCE CENTER CLEARINGHOUSE: A FEASIBILITY STUDY

INTRODUCTION

The feasibility of implementing an information retrieval system for the Technical Assistance Center (TAC) Clearinghouse depends upon the answers provided to two basic questions:

1. Is an information retrieval system necessary and will its advantages outweigh its costs?

2. If an information retrieval system is considered desirable, what particular computer programs (software packages) can be most effectively employed to implement such a system?

I. BACKGROUND INFORMATION

The raison d' etre of the TAC Clearinghouse can be viewed on both a macro- and microcosmic level. Within the microcosm of the Technical Assistance Centers, the need for a centralized TAC Clearinghouse was first enunciated by a panel charged to conduct a performance review of USOE's ESEA Title I Technical Assistane Program during the fall and winter of 1978 (Millman, Paisley, Rogers, Sanders, and Womer, 1979). It was a concern for the "reduction of duplication of efforts across TACs" (p. 60) that led the panel to recommend enhanced techniques of information control:

> The TACs are producing materials (e.g., test reviews, training materials) that have considerable value for Title I evaluation and beyond. By coordinating the collection, storage, and retrieval, updating and distribution of these products, OED could greatly enhance the technical assistancy system...Thus, the panel recommends that printed documents be developed and managed in ways that enhance the quality and cost efficiency of the Technical Assistance Programs. (p. 60)

The rationale behind the establishment of the Clearinghouse on the macrocosmic level lies in the same fundamental problem that has confronted all areas of scholarship since at least the end of World War II, i.e., the information explosion. The proliferation of information -- man's recorded knowledge about himself and his universe -- has increased so dramitically in the twentieth century that the individuals and institutions responsible for information production are now in danger of being submerged in a deluge of their own creation. Information scientists characterize the information explosion in terms of the doubling rate of information, which is variously estimated in the world today at between every three to five years. In other words, the accumulated mass of recorded knowledge is doubling at an alarmingly fast pace. The failure to keep abreast of the information explosion has led in many instances to the costly duplication of research efforts and the under utilization (or nonutilization) of existing information resources.

Thus, the mandate for the establishment of the TAC Clearinghouse was born out of a concern over information control and dissemination as well as the need to avoid duplication of effort. An examination of the options available to establish the TAC Clearinghouse indicates that the use of the computer is the most cost-effective way of producing the card catalog and reproducing it for the other TACs. Consultation with the TAC data processing staff revealed that a number of programs could be used to accomplish this relatively simple operation and that the

American Institutes for Research's (AIR's) contractual arrangements with the Stanford University Computer Center could provide access to the necessary computer hardware.

The question of the feasibility of a computerized information retrieval system for the Clearinghouse can be raised, since the bibliographic information necessary for such a system would already be in machine-readable form. It remains to be determined if such a system is necessary or appropriate, and if desirable, which software package is the most cost-effective to utilize.

II. A COMPUTERIZED, BIBLIOGRAPHIC INFORMATION RETRIEVAL SYSTEM

The computerized information retrieval system in question would exist to fill the information needs of the TACs by providing a range of services that would include, for example, answering SEA and LEA questions, assisting in the preparation of workshops and surveying the literature on specific research topics.

One of the major advantages of implementing a computerized information retrieval system for the TAC Clearinghouse is that the bibliographic data for the system will already exist in machine-readable form. This means that it will not be necessary to engage in any time consuming formatting or coding to get the data into the system. Stated quite simply, the lion's share of the work will have already been performed as a result of the cataloging process. The number of staff hours required to initiate the information retrieval program will consist for the most part of time spent in software development.

Comparison should be made between performing a manual search as opposed to a computer search on the same topic. An information retrieval system has the advantages of speed and volume. Computers are capable of printing output in terms of thousands of lines a minute and dealing with millions of citations at one time. Although the Clearinghouse collection is small at this time, it is expected to grow rapidly as the collection is developed and as new literature on Title I evaluation is produced. For example, thirty papers on Title I were presented at AERA in 1978. In 1979, this number had grown to eighty-six. While physically going through one or two file drawers of research papers may be a feasible manual task now, if the present rate of growth continues, manual searching will soon become impossible.

The most powerful ability of an information retrieval system comes from the performance of multidemensional subject searching. Boolean Logic maintains that all possible relationships between subjects can be expressed by the three local operators OR, AND, and NOT. These operators can be used to combine a string of subjects and retrieve citations that pertain to all the subjects in the string. A manual search through the card catalog on a topic such as "Summer drop-off rates in achievement test scores of neglected and delinquent children in Nebraska" would require the searcher to look under one of the five subject components of this search statement and scan the titles of the articles, hoping to catch references to the other components of this search. The computer simply scans the

citations in the data base to see which ones meet all five criteria and selects only those.

Closely related to multidimensional subject searching is the ability of the computer to combine different fields of the citation to achieve specific results. A field refers to any element of the bibliographic citation that can be used for retrieval such as the author, title, or subject fields. To request citations on "Out-of-level testing workshop materials written after 1978" simply requires combining subject, date, and type-of-material fields. It would not be practical for a human searcher to retrieve citations in this way.

Human searchers can browse the card catalog looking for only citations listed under terms from a controlled vocabulary. The computer is not only capable of doing this but in addition, can scan the titles of articles looking for keywords that may not appear in the controlled vocabulary. It can also scan the abstracts to citations in a like manner when they are present in the data base.

Finally, the commands that are possible within an information retrieval program language offer a variety of powerful information retrieval techniques of which human searchers are not capable. The ability to display thesaurus terms and their relationships on-line, the ability to truncate keywords to retrieve on a variety of word forms for the same concept, and the ability to save a search profile over time for current-awareness purposes are only a few examples.

The question invariably arises as to the costliness of such systems in terms of both computer connect time and staff time spent in program development. Because computers are so widely available, the programs necessary to support such information retrieval systems no longer have to be developed from scratch using such all-purpose languages as COBOL or FORTRAN but can be implemented using program languages specifically designed for information retrieval purposes. In addition, various private corporations offer bibliographic data base maintenance services for private files.

The costs of computer operations have been steadily decreasing as technological innovations increase. Although the literature on the subject has scarcely kept pace with the innovations, numerous studies in the library and information science literature have accented the decrease in costs due to library automation. Kilgour (1972) found that the cost of a computer-produced library card at the Ohio College Library Center was $0.106, while the cost of manually producing a

card was $0.236. Bregzis (1973) reports similar savings in a study of the conversion of bibliographic records to machine-readable form.

After an examination of the available options, it is apparent that an information retrieval system is both appropriate and cost-effective.

III. SOFTWARE PACKAGES

If the decision is made to implement an information retrieval system for the TAC Clearinghouse, it must be still decided what software package to utilize for this purpose. Basically, this decision involves whether or not to develop the software in-house at AIR or to contract for the use of a commercially available package. Evaluation of software packages for information retrieval purposes can be based on four criteria, according to Wos (1969):

> Simplicity: the system must be easy to learn, install, and use.
>
> Flexibility: the system must be able to perform a wide variety of functions.
>
> Comprehensiveness: the system must be able to perform specialized functions for a given environment.
>
> Efficiency: the system must reduce programming time and effort without greatly increasing machine time.

To these four criteria can be added a fifth, namely:

> Cost: the system must compare favorably in terms of cost with other competing systems.

The remainder of this chapter compares a software package available in-house at AIR with a locally avaliable commercial product on the basis of these five criteria.

Lockheed's DIALOG Private File Service

Lockheed's DIALOG division provides nationwide access to almost all of the major bibliographic data bases in the natural and social sciences. Of particular interest to the TAC effort are the ERIC and Psychological Abstracts data bases. Users of the DIALOG service can also elect to have their own private files put onto DIALOG, with passwood control to maintain the confidentiality of the files. Bibliographic information must be provided on magnetic tape to DIALOG in a specific format. The costs of the system are as follows:

> Access: $35.00 an hour computer connect time.
>
> Offline Prints: $.10 per citation.
>
> Storage: $12.00 a month.
>
> Updates: $50.00 minimum charge per update.

Average Search Cost: (based on 15 minutes connect time and 25 citations retrieved): $11.25.

The advantages of employing the DIALOG software system lie in the powerful retrieval capabilities of the language itself. Any DIALOG-trained user would be able to access the system readily while the potential exists for other TACs with DIALOG contacts to access the system as well. Training is supplied by Lockheed to all contractual users. The only pieces of equipment required are a telephone and a computer terminal. No programming would be involved in initiating the system.

The major disadvantages with DIALOG lie in the costs for connect time and update procedures. The connect costs do not compare favorably with the in-house system under consideration, while the update fees are so excessive as to preclude frequent updating of the system. Most importantly, the bibliographic data tape to be submitted to DIALOG would be in a different format from that necessary to produce the catalog cards. This means, in effect, that all Clearinghouse materials would have to be processed twice: once for catalog card generation and once for entry into the information retrieval system. A software package that could avoid this costly duplication of effort would be much more desirable.

The Clearinghouse Information Retrievel System available through DIALOG would be accessible to the TACs only after they had completed contractual arrangements with Lockheed. As already mentioned, training is provided by Lockheed. A telephone and computer terminal are the only pieces of equipment necessary to access the system. Communication links with the DIALOG computer in Palo Alto, California, are provided through either TYMSHARE or TELENET Corporations at a cost that is greatly reduced compared with standard commercial telecommunication charges. Contractual arrangements would have to be made with these firms by each of the TACs.

The Stanford Public Information Retrieval System

The Stanford Public Information Retrieval System (SPIRES) was developed to meet the need for a generalized data base management system. Originally developed as a software package for an automation project at the Stanford University Libraries, SPIRES has evolved into a comprehensive information retrieval program that serves as an "extended library and offers a favorable cost-benefit

ratio" (Ferguson, 1970, p. 15). As a result of its original association with library catalog automation endeavors, SPIRES is capable of printing catalog cards as well as functioning as an information retrieval system. SPIRES is currently used by the Research Library Information Network (RLIN) for just this purpose. The costs of the system are as follows:

Access: $6.00 an hour computer connect time.

Offline Prints: Approximately $.03 each.

Storage: $12.50 a month.

Updates: No minimum charge; cost of updating the file consists only of access charges.

Average Search Cost: (based on 15 minutes connect time and 25 citations retrieved): $2.10.

Compared with the DIALOG system, SPIRES computer connect costs are minimal, and there are no update charges. Thus, frequent updates can be performed on the data base to ensure its operation at maximal capacity. By far the most important advantage to SPIRES lies in the fact that the language can be used to produce both the catalog cards and the information retrieval system. Thus, items to be added to the Clearinghouse will require one-step processing. The programming effort required to initiate the SPIRES retrieval system will be greater than that for DIALOG, but, since the same programming language and record formats can be used for card catalog and information retrieval production, the overall effort should be less than that to program and format both tasks separately.

The disadvantage of SPIRES is that a greater intellectual effort will be required to set up the system initially in terms of file definition, formatting, and programming. Once established, however, the system will employ a program language that is user-oriented and will require a minimum of training to access and update.

The Clearinghouse Information Retrieval System available through SPIRES would be directly accessible to the TACs at Stanford University using AIR's password. Training on the system would be provided by the Clearinghouse staff at a seminar attended by representatives from each TAC. As with DIALOG, a telephone and a computer terminal are the only pieces of equipment needed to access the system. Inexpensive telecommunication hook-ups can be provided by either TYMSHARE or TELENET through contractual arrangements with each TAC.

IV. EVALUATION

Evaluation of these two software packages can be based on the five criteria outlined earlier:

- Simplicity: Both DIALOG and SPIRES are relatively simple to learn and use. Installation of SPIRES, however, will require a greater level of technical expertise.

- Flexibility: DIALOG essentially offers no flexibility in the variety of functions it performs. Once the records are entered into the data base, methods of access and output are controlled by Lockheed. SPIRES, on the other hand, offers a high degree of flexibility, since all decisions on format and input-output variables are dependent upon programmer specifications.

- Comprehensiveness: Both DIALOG and SPIRES are able to perform a variety of complex information retrieval functions. Both systems provide the techniques necessary to support an information retrieval system, such as that envisioned for the Clearinghouse.

- Efficiency: It is inappropriate to consider the ratio of programming time to machine time with reference to DIALOG. As far as SPIRES is concerned, this ratio might not be the most cost-effective one in existence, but at the low connect cost of $6 an hour, the issue can hardly be considered crucial in an overall evaluation of the system.

- Cost: A comparison of costs is given in Table 1. SPIRES is a far less expensive system to access and update than DIALOG based on the comparison of one or all of the cost variables. Cost of implementing SPIRES will be greater in terms of TAC staff time required, but these monies have already been budgeted for the task. Since programming for the information retrieval system and card catalog production aspects of the task can be consolidated under SPIRES, the amount of programming effort that will be expended on the Clearinghouse overall should level off at an acceptable figure.

Criteria that can be used to evaluate the effectiveness of an information retrieval system overall, suggest Lancastar and Climenson (1968), are as follows:

- Coverage: the proportion of the total useful literature on a particular topic that has been collected and indexed into the system.

- Usability: the value of the literature on a particular topic, in terms of such criteria as age, reliability, and comprehensibility.

Table 1

Cost Comparisons for Two Information Retrieval Systems

Cost Catagories	SYSTEM	
	SPIRES	DIALOG
ACCESS	$ 6.00/hour	$35.00/hour
OFFLINE PRINTS	$ .03/citation	$ .10/citation
STORAGE	$12.50/month	$12.00/month
UPDATES	0	$50.00/update
AVERAGE SEARCH	$ 2.10	$11.25

- Recall: the ability of the system to retrieve wanted documents.
- Precision: the ability of the system to screen out unwanted documents.
- Response time: the average time required to obtain a satisfactory response from the system.
- Presentation: the form in which search results are presented.
- User effort: the amount of effort the user must personnally expend in order to consult the system.

The proposed Clearinghouse information retrieval system, if implemented, will be evaluated according to these criteria.

V. CONCLUSIONS

The services offered by the Clearinghouse should be typified in terms of the notion of information utility, as discussed by Seracevic (1974) among others. The notion combines the concept of information as a commodity and the concept of a utility as a public service oriented towards providing commodities in the public interest. The idea is to provide information services that are widely available in such a way that the potential users use them at their convenience, when needed, and for their individual purposes. The basic purpose of an information utility is to increase the availability of information and broaden the population of users. The need for an information utility arises in answer to the fundamental problem discussed at the beginning of this paper: the information explosion. Modern information problems are universal and global: an overabundance of information at the source and scarcity of relevant information for the user at the destination. The individuals involved in Title I evaluation are not less caught up in this quandry than are other researchers in the natural or social sciences. The growth rate and the coverage of Title I evaluation documentation are still within manageable limits. The cost and the effort to be expended on control are minimal when compared with the benefits to be reaped.

Based on the results of this study, it seems advisable to recommend SPIRES as the software package for the proposed Clearinghouse information retrieval system for three reasons:

- Costs will be considerably lower than those involved in utilizing a commercially available service.
- Single rather than dual step processing of Clearinghouse materials will be possible using SPIRES with regard to the tasks of card catalog production and information retrieval system maintenance.
- All aspects of the data base's management will be centralized in the Clearinghouse.

An examination of the available options suggests that a computerized information retrieval system is both a viable and cost-effective component in the establishment of a TAC Clearinghouse.

REFERENCES

Bregzis, R. Some cost aspects of on-line input of bibliographic records. In F. Slater (Ed.), Cost reduction for special libraries and information centers. Washington, DC: American Society for Information Science, 1973.

Ferguson, D. Information retrieval (SPIRES) and library automation (BALLOTS) at Stanford University. Stanford, CA: Stanford University Computer Center, 1970. (ERIC No. ED 044 049.)

Kilgour, F. G. Evolving computerizing, personalizing. American Libraries, 1972, 3(2), 141-147.

Lancaster, R. W., & Climenson, W. D. Evaluating the economic efficiency of a document retrieval system. Journal of Documentation, 1968, 24(1), 16-40.

Millman, J., Paisley, W., Rogers, W. T., Sanders, J. R., & Womer, F. B. Performance review of USOE's ESEA Title I evaluation technical assistance program. Washington, DC: Hope Associates, 1979.

Seracevic, T. Questions about the notion of information utilities. American Society of Information Science Proceedings, 1974, 11, 1-7.

Wos, C. M. The evaluation of file-management software packages. American Society of Information Science Proceedings, 1969, 6, 215-222.